JPRS: 4687

12 June 1961

THE ORGANIZED SEARCH PRINCIPLE IN SYSTEMS WITH AUTOMATIC OPTIMIZATION

By I. M. Gel'fand and M. L. Tsetlin

-USSR-

Distributed by:

OFFICE OF TECHNICAL SERVICES
U. S. DEPARTMENT OF COMMERCE
WASHINGTON 25, D. C.

U. S. JOINT PUBLICATIONS RESEARCH SERVICE
1636 CONNECTICUT AVE., N.W.
WASHINGTON 25, D. C.

JPRS: 4687

CSO: 1846-S

THE ORGANIZED SEARCH PRINCIPLE IN SYSTEMS WITH AUTOMATIC OPTIMIZATION

[Following is a translation of the article by, Corresponding Member of the Academy of Sciences USSR, I. M. Gel'fand, and M. L. Tsetlin, in Doklady Akademii Nauk SSR (Reports of the Academy of Sciences USSR), Vol 137, No 2, Moscow, 1961, pages 295-298].

In many practical problems need arises to control the complex systems with high number of degrees of freedom. The study of physiological processes, such as construction of motion, analysis of afferent nerves etc., leads to this category of problems. The attempts to solve these problems by means of classical mathematics very often turned out to be unsuccessful. Even if there is no difficulties to find an appropriate algorithm to solve all cases of a problem, the realization of this algorithm is not always feasible due to the gap between speed of solution required by a problem and offered by contemporary computing technique in real time domain.

In situations like these the solution may be achieved by the use of "Organization", which is a familiar pattern in physiology, particularly in human behavior. In other words the solution is attained at the expense of avoiding to consider the most laborious chaotic situations. Let us note, that in many occasions these disadvantageous chaotic situations are most probable cases in a sense of formal mathematical treatment.

This paper is an essay on the solution of a problem in automatic control, namely a problem of automatic optimization with a high number of operative parameters.

It is well known (Ref 1 and 2) that an adaptive control system is an automatic feedback control system which continuously adapts itself to a changing environment by monitoring its own behavior and automatically adjusting its parameters toward optimum performance. We will call these adaptive or self-adptive systems simply systems with automatic optimization. Subsequently we will present the automatic optimization principle based upon a special method of organized-non-local search, suggested by I. M. Gel'fand. This method has turned out to be a very effective one in computing of an extremum (minimum or maximum).

Now, let $F(x_1, \ldots, x_n, y_1, \ldots, y_m)$ be the output function of our system with automatic optimization. We will call a group of independent variables $x_1, \ldots, x_n$ an operative group of arguments. The values of these arguments vary in the process of automatic search. The arguments $y_1, \ldots, y_m$ are implicit system parameters, generally are functions of time, therefore the output function may be written in this form $F(x_1, \ldots, x_n, y_1, \ldots, y_m) = \Phi(x_1, \ldots, x_n, t)$.

We have to note, that in self-adaptive control systems the output function $\Phi(x_1, \ldots, x_n, t)$ usually is not given in analytical or any other concrete form, so that the selection of necessary optimizing values of operative parameters is an experimental process. The functional dependance of Φ of time, which is not known explicitly, leads to necessity of continuous search for desired values of arguments. Consequently the speed of search is a very important factor, i.e. roughly speaking, the correction of the output must be accomplished fast enough to catch up with the function before it is too late to affect it.

In other words, in order to keep the optimized output function in certain desired limits, it is necessary to reach these satisfactory values before the function undergoes a considerable change in its value. The proximity of instantaneous value of the function to its extremum could not be used as a criterion of the system performance; a following functional can be used instead

$$T^{-1}\int_0^T \Psi(\Phi(x_1, \ldots, x_n, t))\, dt. \qquad (1)$$

The function Ψ may be chosen in different ways. Thus, for instance, with $\Psi(\Phi) = \Phi$ the value of this functional corresponds to so called "price of the search", which is found in the simplest systems of automatic minimization with single operative parameter and single minimum of function Φ (Ref 1).

In the problems where it is sufficient to maintain the values of Φ no greater than a certain prescribed level C (Minimization on the level) it will be convenient to use a criterion obtained from (1) with the assumption

$$\Psi(\Phi) = 0 \text{ with } \Phi \leqslant C; \qquad \Psi(\Phi) = 1 \text{ with } \Phi > C. \qquad (2)$$

The automatic search may by accomplished by means of different methods, which could be divided into three groups.

First group is the method of random search, in which the experimental trials of necessary parameters are performed independently one from another. In these cases either all points of the operative parameters space is scanned in a certain order; or these points will be chosen at random and not changed as long as values of Φ are still satisfactory (homeostat principle, Ref 3 and 4). During continuous search by these methods, values of Φ are not improving systematically from trial to trial, hence "the price of the search" is high.

The second group is a method of local automatic search, which

provides the analysis of results of each trial in order to determine the values for the next try. To this group belong methods like gradients, relaxation, most rapid descent and some others. Their common feature is localism: an operative point continuously moves in the space of operative arguments and the preparation of the next trial is performed taking into account values of function Φ from the previous trial and expanding them in their neighbourhood. Local search method provides the better values of Φ in the process of search, which makes this method superior to the method of random search.

However, using only local search, i.e. in the nearest neighbourhood, this method only slightly utilizes the specificity of the optimized function. With small gradients all local methods are not efficient, forcing one to resort to multiple change of the operative step in the direction of motion, which results in the considerable slowing down of search*. (* Besides there is always danger of recycling of the search around some insignificant "shallow sink").

The systems of automatic optimization with local method of search have been described in detail in works of A. A. Feldbaum (Ref 2, 5, and 6), where several schematic diagrams of these systems are given.

The third group is a method of non-local search. The characteristic feature of this group is discontinuity of motion of the operative point in the parametric space. This sharply increases volume of space scanned per unit time; as there arises possibility to utilize the specificity of the construction of the function Φ; the process of optimization is considerably faster.

The simplest non-local method is a combination of local and homeostat methods. This method is often used in computational practice and is as follows: At first a point at random is chosen then from it the descent is made in accordance with a preferred local method until the change of the function becomes small. After this, another point at random is chosen and descent is made etc. This is, for instance, the non-local gradient method.

This method of non-local search possesses a demerit that after each descent the value of Φ again (and, generally speaking, considerably) increases; the information about function gained during a local descent subsequently is not utilized. On account of the systematic exit into the region of large values of and consequently the necessity of protracted use of local methods, "the price of the search" is relatively high.

We will now describe our method of non-local search, which we will call a method of the ravines. This method implies that operative parameters $x_1, \ldots, x_n$ could be divided into two groups. The first group includes almost all the parameters and variations of those parameters results in considerable change of function Φ. Hence, the tuning on these parameters is simple and rapid. These parameters we will call unimportant. The second group of variables includes small (may be 2 or 3) number of functions of $x_1, \ldots, x_n$ variations of which results in relatively small changes of Φ. We will call them essential variables. Naturally, the division of parameters into two groups is

the time dependent and therefore must be accomplished automatically. This division of parameters, of course, is impossible to obtain for all mathematical functions. However for all practical cases (problems of applied physics, of engineering etc.) such division, apparently, is always possible to accomplish. Realizing the difficulty of exact definition of these concepts, we will nevertheless risk to call them here well organized functions.

Automatic search then is going in the following way. At first an arbitrary point X_0 (Fig 1) is chosen. From this point a descent is made by the gradient (it is possible, of course, to use any other local method). This descent better be a course one; if, for instance the next step will decrease value of Φ less than 5% to 15%, then it should be stopped. The fact is that as soon as the descent by gradient ceases to have substantial effect on the value of Φ, we get into a zone, where variables of both groups become equivalent; and without any noticeable progress due to essential variables we will continue to wander irregularly varying the unimportant variables. As a matter of fact, this is the cause of low efficiency of the local methods.

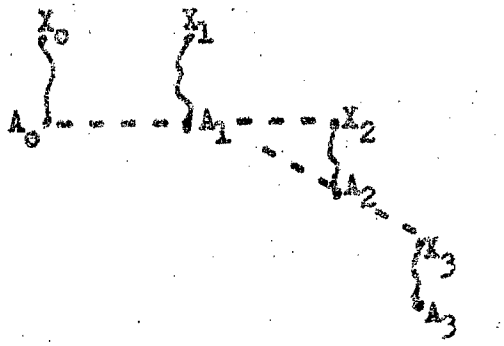

Fig 1

Let us assume that the gradient descent brought us from point X_0 to point A_0. Subsequently another point X_1 is selected in the neighbourhood of X_0 but at a distance substantially greater than the gradient-step of descent. From the point X_1 the gradient descent is made. Now we have two points A_0 and A_1 in a ravine and we will make the next step along the ravine. The points A_0 and A_1 we will connect by a straight line, extend the line beyond point A_1 and select a new point X_2 on it at a distance from point A_1, which we will call the ravine-step. This chosen step for well organized functions is considerably greater than a gradient-step. The selection of the ravine-step is done experimentally by trials and it is an important characteristic of function Φ. From point X_2 again the gradient descent is made and we get a point A_2. Point X_3 is found on the straight line extended beyond line A_1-A_2 in a similar manner as the line A_0-A_1 was used to find a point X_2, and the whole process is repeated. Points X_1 are chosen in the area where small values of Φ are expected* nevertheless the whole search is performed principally in a zone of small values of Φ (* However, at the exact locations of points X_1 the values of Φ may not be small, due to the effect of unimportant parameters). With a reasonable choice of a ravine-step the whole process of search adapts itself to the course along the ravine, so that values of gradient descents become much smaller than a value of a ravine-step** . (** The selection of a ravine-step determines qualitative characteristics of tactics of the search. With a chosen ravine-step we "go over

the hills and round the high mountains". The ravine-step serves as a scale factor). As a result "the price of search" is considerably smaller than in local methods or previously described primitive non-local method. Thus, with eight or ten variable parameters in the problems of phase analysis, "the price of search" will be hundreds of times lower.

We will note, that for the selection of ravine-step it would be helpful to use the functional (1) with Ψ function chosen in accordance with (2). The length of the ravine-step corresponds to a certain fixed value of this functional. If the ravine-step is too small the functional becomes small and search goes slow. If the ravine-step is very large the value of the functional increases and the search is course.

The selection of a gradient for the functions of many variables is tied with the necessity of fixing $(n + 1)$ values of Φ and becomes too cumbersome and tedious. The following method may be used here. The certain initial probability values $p_1, \ldots, p_n$ is given to the variables $x_1, \ldots, x_n$. Subsequently in accordance with probability values few directions are selected and a partial gradient is computed. Depending on the results of motion by a partial gradient the initial distribution of probabilities is corrected, so that the adaptation of direction of motion toward direction of total gradient* takes place. (* These considerations are closely related to the behavior of automata, which would be a topic for discussion in a separate M. L. Tsetlin paper). It has been mentioned above, instead of gradients, any trial of local situations is possible to use.

It is important to note, that for functions changing rapidly with time, the correlation between two points separated by a ravine-step during the shift along the ravine becomes small; and therefore the ravine-method in this sense is approaching the non-local gradient method. Further increase of dependance of Φ of time leads to decrease of local correlation and this case approaches the random method. In substance the ravine-method includes in it non-local gradient method and random search and yields essential profit in the cases when the speed (frequency) of search is greater than rate of change (frequency) of function Φ; in all other cases this method is little better than gradient and random methods. A similar situation exists also when the degree of organization of function is decreased.

We will note in conclusion, that it seems probable to us the existence of connection (analogy) between primitive follow-up systems, methods of random search, local and non-local methods of automatic optimization (adaptive systems), on the one hand, and construction of motion of human being and animals, on the other hand. This thought had been discussed first by N. A. Bernshteyn (Ref. 7).

We wish to express our acknowledgement to M. A. Yevgrafov, L. N. Ivanova, and I. I. Pyatetskiy-Shapiro for their numerous and beneficial exchanges of opinion.

Received 14 December 1960.

REFERENCES

1. CH'IEN, Hsueh-sen, Tekhnicheskaya kibernetika (Technical Cybernetics), Foreign Literature Publishing House, Moscow, 1956.

2. FEL'DBAUM, A. A. Vychislitel'nye ustroystva v avtomaticheskikh sistemakh (Computers in Control Systems), Moscow, 1959.

3. ASHBURY, W. R., An Introduction to Cybernetics, London, 1956.

4. SAVINOV, G. V., Problemy kibernetiki (Problems of Cybernetics), Vol 4, Moscow, 1960.

5. FEL'DBAUM, A. A., Avtomatika i telemekhanika (Automatics and Telemechanics), 17, No 11, 1956.

6. FEL'DBAUM, A. A., op. cit., 19, No 8, 1958.

7. BERNSHTEYN, N. A., O postroyenii dvizheniy (On Construction of Motion), Moscow, 1948.

10,443

- END -